Joint source-channel coding with feedback

Victoria Kostina

California Institute of Technology

Pasadena, CA 91125

vkostina@caltech.edu

Yury Polyanskiy

Massachusetts Institute of Technology

Cambridge, MA 02139

yp@mit.edu

Sergio Verdú

Princeton University

Princeton, NJ 08544

verdu@princeton.edu

Abstract

This paper quantifies the fundamental limits of variable-length transmission of a general (possibly analog) source over a memoryless channel with noiseless feedback, under a distortion constraint. We consider excess distortion, average distortion and guaranteed distortion (d-semifaithful codes). In contrast to the asymptotic fundamental limit, a general conclusion is that allowing variable-length codes and feedback leads to a sizable improvement in the fundamental delay-distortion tradeoff. In addition, we investigate the minimum energy required to reproduce k source samples with a given fidelity after transmission over a memoryless Gaussian channel, and we show that the required minimum energy is reduced with feedback and an average (rather than maximal) power constraint.

Index Terms

Variable-length coding, joint source-channel coding, lossy compression, single-shot method, finite-blocklength regime, rate-distortion theory, feedback, memoryless channels, Gaussian channels, energy-distortion tradeoff, Shannon theory.

I. INTRODUCTION

A famous result of Shannon [1] states that feedback cannot increase the capacity of memoryless channels. Burnashev [2] showed that feedback improves the error exponent in a variable-length setting. Polyanskiy et al. [3] showed that allowing variable-length coding and non-vanishing error-probability ϵ boosts the ϵ-capacity of the discrete memoryless channel (DMC) by a factor

of $1 - \epsilon$. Furthermore, as shown in [3], if both feedback and variable-length coding are allowed, then the asymptotic limit $\frac{C}{1-\epsilon}$ is approached at a fast speed $O\left(\frac{\log \ell}{\ell}\right)$ as the average allowable delay ℓ increases:[1]

$$(1 - \epsilon) \log M^\star(\ell, \epsilon) = \ell C + O\left(\log \ell\right) \tag{1}$$

where $M^\star(\ell, \epsilon)$ is the maximum number of messages that can be distinguished with error probability ϵ at average delay ℓ, and C is the channel capacity. This is in contrast to channel coding at fixed blocklength n where in most cases only a $O\left(\frac{1}{\sqrt{n}}\right)$ convergence rate is attainable, even when feedback is available, see [3], [4]. Thus, variable-length coding with feedback (VLF) not only boosts the ϵ-capacity of the channel, but also markedly accelerates the speed of approach to it. Moreover, zero-error communication is possible at an average rate arbitrarily close to capacity via variable-length coding with feedback and termination (VLFT) codes, a class of codes that employs a special termination symbol to signal the end of transmission, which is always recognized error-free by the receiver [3]. As discussed in [3], the availability of zero-error termination symbols models that common situation in which timing information is managed by a higher layer whose reliability is much higher than that of the payload.

In [5], we treated variable-length data compression with nonzero excess distortion probability. In particular, we showed that in fixed-to-variable-length compression of a block of k i.i.d. source outcomes, the minimum average encoded length $\ell^\star(k, d, \epsilon)$ compatible with probability ϵ of exceeding distortion threshold d satisfies, under regularity assumptions,

$$\ell^\star(k, d, \epsilon) = (1 - \epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}} e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O\left(\log k\right) \tag{2}$$

where $R(d)$ and $\mathcal{V}(d)$ are the rate-distortion and the rate-dispersion functions, and Q is the standard normal complementary cumulative distribution function. The second term in the expansion (2) becomes more natural if one notices that for $Z \sim \mathcal{N}(0, 1)$,

$$\mathbb{E}\left[Z \cdot 1\left\{Z > Q^{-1}(\epsilon)\right\}\right] = \frac{1}{\sqrt{2\pi}} e^{-\frac{Q^{-1}(\epsilon)^2}{2}}. \tag{3}$$

As elaborated in [5], the expansion (2) has an unusual feature: the asymptotic fundamental limit is approached from the "wrong" side, e.g. larger dispersions and shorter blocklengths reduce the average compression rate.

---

[1]Unless explicitly noted, all $\log$ and $\exp$ in this paper are to arbitrary matching base, which also defines units of entropy, information density and mutual information.

In this paper, we consider variable-length transmission of a general (possibly analog) source over a DMC with feedback, under a distortion constraint. This variable-length joint source-channel coding (JSCC) setting can be viewed as a generalization of the setups in [3], [5], which, as explained above, analyze the problem without source coding and channel transmission, respectively. Related work includes an assessment of nonasymptotic fundamental limits of fixed-length JSCC in [6]–[8], a dynamic programming formulation of zero-delay JSCC with feedback in [9], and a practical variable-length almost lossless joint compression/transmission scheme in [10]. Various feedback coding strategies are discussed in [11]–[17]. Practical feedback communication schemes in the literature that implement VLF include [18]–[22].

We treat several scenarios that differ in how the distortion is evaluated and whether a termination symbol is allowed. In all cases, we analyze the *average delay* required to achieve the objective. The results in Section III, where as before, $C, R(d), \mathcal{V}(d)$ denote the channel capacity, and the source rate-distortion and rate-dispersion functions, respectively, are summarized as follows:

- Under the *average distortion* criterion, $\mathbb{E}\left[\mathsf{d}(S^k, \hat{S}^k)\right] \leq d$, the minimal average delay $\ell^\star(k, d)$ attainable via VLF codes transmitting k source symbols satisfies

$$\ell^\star(k, d)C = kR(d) + O\left(\log k\right). \tag{4}$$

- Under the *excess distortion probability* criterion, $\mathbb{P}[\mathsf{d}(S^k, \hat{S}^k) > d] \leq \epsilon$, the minimal average delay attainable via VLF codes transmitting k source symbols satisfies

$$\ell^\star(k, d, \epsilon)\, C = (1 - \epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}}e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O\left(\log k\right). \tag{5}$$

- Under the *guaranteed maximal distortion* criterion, $\mathbb{P}[\mathsf{d}(S^k, \hat{S}^k) > d] = 0$, the minimal average delay attainable via VLFT codes transmitting k source symbols satisfies

$$\ell_t^\star(k, d, 0)\, C = kR(d) + O\left(\log k\right). \tag{6}$$

Similar to (1), approaching the limits in (4), (5) and (6) only requires an extremely thin feedback link, namely, the decoder sends just a single acknowledgement signal once it is ready to decode (stop-feedback)[2]. Note that (5) exhibits significant similarities with (2): the

---

[2]Stop-feedback is not to be confused with the termination symbol, which is a special symbol that the encoder can transmit error-free in order to tell the decoder that the transmission has ended and it is time to decode.

asymptotic limit is approached from below, i.e. in contrast to the results in [6], [23], [24], smaller blocklengths and larger source dispersions are beneficial. Note also that the first term of the expansion in (5) can be attained with variable-length codes without feedback.

Interestingly, naive separated source/channel coding fails to attain any of the limits mentioned. For example, it approaches the asymptotic fundamental limit from above, e.g. even the sign of the second term in (5) is not attainable. This observation led us to believe, initially, that competitive schemes in this setting should be of successive refinement and adaptation sort such as in [25], [26], or dynamic programming-like as in [9], [27]. It turns out, however, that like the fixed-length JSCC achievability schemes in [6], [7], attaining limits (4)-(6) requires a rather simple variation on the separation architecture: one only needs to allow a variable-length interface between the source coder and the channel coder. Note that typically, separation is understood in the sense that the output of the source coder (compressor) is treated as pure bits, which can be arbitrarily permuted without affecting performance of the concatenated scheme [8], [28]. Similarly, the performance of a *variable-length separated scheme* is insensitive to permutations (but not additions or deletions) of the bits at the output of the source coder. These semi-joint achievability schemes are the subject of Section II. They form the basis for the lossy joint source-channel codes, which are the subject of Section III.

Energy-limited codes are the subject of Section IV. The optimal energy-distortion tradeoff achievable in the transmission of a general source over the AWGN channel is studied in Section V. In that setting, disposing of the restriction on the number of channel uses per source sample, we limit the total available transmitter energy E and we study the tradeoff between the source dimension k, the total energy E and the fidelity of reproduction. Related prior work includes a study of asymptotic energy-distortion tradeoffs [29] and a nonasymptotic analysis of energy per bit required to reliably send k bits through an AWGN channel [30]. The main results in Section V are the following:

- Under *average distortion* constraint, the total minimum energy required to transmit k source samples over an AWGN channel *with feedback* satisfies

$$E_f^\star(k, d) \cdot \frac{\log e}{N_0} = kR(d) + O(\log k),$$ (7)

  where $\frac{N_0}{2}$ is the noise power per degree of freedom.

- Under *excess distortion probability* constraint, the minimum energy required to transmit k

source samples over an AWGN channel *without feedback* satisfies

$$E^\star(k, d, \epsilon) \cdot \frac{\log e}{N_0} = kR(d) + \sqrt{k\left(2R(d)\log e + \mathcal{V}(d)\right)}Q^{-1}(\epsilon) + O(\log k) \qquad (8)$$

- Under *excess distortion probability* constraint, the total minimum average energy[3] required to transmit k source samples over an AWGN channel *with feedback* satisfies

$$E_f^\star(k, d, \epsilon) \cdot \frac{\log e}{N_0} = kR(d)(1 - \epsilon) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}}e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O(\log k) \qquad (9)$$

Like (5), particularizing (9) to $\epsilon = 0$ also covers the case of guaranteed distortion. The first term in the expansion (9) can be achieved even without feedback, as long as $\epsilon > 0$ and the power constraint is understood on the average over the codebook.

We point out the following parallels between the variable-length codes and energy-limited-codes.

- Under average distortion, in both cases the fundamental limit is approached at speed $O\left(\frac{\log k}{k}\right)$ (cf. (4), (7)).

- Allowing a non-vanishing excess-distortion probability and variable length (or variable power) boosts the asymptotic fundamental limit by a factor of $1 - \epsilon$.

- Allowing both feedback and variable length (or power) leads to expansions (5), (9), in which shorter blocklengths are beneficial.

- As long as feedback is available, in both variable length coding with termination and average energy-limited coding, guaranteed distortion ($\epsilon = 0$) can be attained, even though the channel is noisy.

II. FEEDBACK CODES FOR NON-EQUIPROBABLE MESSAGES

In this section we consider joint source-channel coding assessing reliability by the probability that the (possibly non-equiprobable) message is reproduced correctly. These results lay the foundation for the analysis of joint source-channel coding under distortion constraints presented in Section III. Our key tool will be two extensions of the channel coding bounds for the DMC with feedback from [3]. VLF and VLFT codes are formally defined as follows.

---

[3]The energy constraint in (8) is understood on a per-codeword basis. The energy constraint in (9) is understood as average over the source and noise realizations.

Definition 1. *A variable-length feedback code (VLF) transmitting message W (taking values in $\mathcal{W}$) over the channel $\{P_{Y_i|X^iY^{i-1}}\}_{i=1}^{\infty}$ with input/output alphabets $\mathcal{A}/\mathcal{B}$ is defined by:*

1) *A random variable $U \in \mathcal{U}$ revealed to the encoder and decoder before the start of the transmission.*

2) *A sequence of encoding functions $\mathsf{f}_n\colon \mathcal{U} \times \mathcal{W} \times \mathcal{B}^{n-1} \mapsto \mathcal{A}$, specifying the channel inputs*

$$X_n = \mathsf{f}_n\left(U, W, Y^{n-1}\right). \tag{10}$$

3) *A sequence of decoding functions $\mathsf{g}_n\colon \mathcal{U} \times \mathcal{B}^n \mapsto \mathcal{W}$, $n = 1, 2, \ldots$*

4) *A non-negative integer-valued random variable τ, a stopping time of the filtration $\mathcal{F}_n = \sigma\{U, Y_1, \ldots, Y_n\}$.*

The final decision $\widehat{W}$ is computed at the time instant τ

$$\widehat{W} = \mathsf{g}_\tau(U, Y^\tau) \tag{11}$$

The value $\mathbb{E}\left[\tau\right]$ is the average transmission length of the given code.

A very similar concept is that of an VLFT code:

Definition 2. *A variable-length feedback code with termination (VLFT) transmitting $W \in \mathcal{W}$ over the channel $\{P_{Y_i|X^iY^{i-1}}\}_{i=1}^{\infty}$ with input/output alphabets $\mathcal{A}/\mathcal{B}$ is defined similarly to VLF codes with the exception that condition 4) in Definition 1 is replaced by*

4') *A non-negative integer-valued random variable τ, a stopping time of the filtration $\mathcal{G}_n = \sigma\{W, U, Y_1, \ldots, Y_n\}$.*

The idea of allowing τ to depend on the true message W models the practical scenarios (see [3]) where there is a highly reliable control layer operating in parallel with the data channel, which notifies the decoder when it is time to make a decision.

The two special cases of the above definitions are stop-feedback and fixed-to-variable codes:

1) *stop-feedback codes* are a special case of VLF codes where the encoder functions $\{f_n\}_{n=1}^{\infty}$ satisfy:

$$f_n(U, W, Y^{n-1}) = f_n(U, W). \tag{12}$$

Such codes require very limited communication over feedback: only a single signal to stop the transmission once the decoder is ready to decode.

2) *fixed-to-variable codes*, defined in [31], are also required to satisfy (12), while the stopping time is[4]

$$\tau = \inf\{n \geq 1 : g_n(U, Y^n) = W\}, \tag{13}$$

and therefore, such codes are zero-error VLFT codes.

For both VLF and VLFT codes, we say that a code that satisfies $\mathbb{E}[\tau] \leq \ell$ and $\mathbb{P}\left[W \neq \widehat{W}\right] \leq \epsilon$, when averaged over U, message and channel, is an (ℓ, ϵ) code for source/channel $\left(W, \{P_{Y_i|X^iY^{i-1}}\}_{i=1}^{\infty}\right)$.

The random variable U serves as the common randomness shared by both transmitter and receiver, which is used to initialize the codebook. As a consequence of Caratheodory's theorem, the amount of this common randomness can always be reduced to just a few bits: as shown in [3, Theorem 19], if there exists an (ℓ, ϵ) code with $|\mathcal{U}| = \infty$, then there exists an (ℓ, ϵ) code with $|\mathcal{U}| \leq 3$. Allowing for common randomness does not affect the asymptotic expansions, but leads to more concise expressions for our non-asymptotic achievability bounds. Furthermore, no common randomness is needed at all if the channel is symmetric [3, Theorem 3].

First, we quote an achievability result [3, (107)-(118)]. Let P_Y be the capacity achieving output distribution of the DMC. Denote information density as usual:

$$\imath_{X;Y}(a; b) \triangleq \log \frac{P_{Y|X}(b|a)}{P_Y(b)} \tag{14}$$

Theorem 1 ([3]). *For every DMC with capacity C, any M and probability of error ϵ there exists an (ℓ, ϵ) stop-feedback code for the message W taking M values[5] such that*

$$C\ell \leq \log M + \log \frac{1}{\epsilon} + a_0,$$

where

$$a_0 \triangleq \max_{x,y} \imath_{X;Y}(x; y). \tag{15}$$

We note that a similar result is also contained in many other works, starting from [2] and later in [32], [33].

---

[4]As explained in [31], this model encompasses fountain codes in which the decoder can get a highly reliable estimate of τ autonomously without the need for a termination symbol.

[5]Although the result in [3] is stated for average (over equiprobable messages) error probability, it in fact applies to arbitrary distribution of messages, so W need not be equiprobable. Furthermore, if we do not insist on $|\mathcal{U}| \leq 3$, Theorem 1 even applies to maximal probability of error.

Next, we tighten Theorem 1 in the case of non-equiprobable messages.

Theorem 2. *For every DMC with capacity C and random variable W there exists an (ℓ, ϵ) stop-feedback code for W with*

$$C\ell \leq H(W) + \log \frac{1}{\epsilon} + a_0 \tag{16}$$

where a_0 is defined in (15).

Proof: If $H(W) = \infty$ then there is nothing to prove, so we assume otherwise. Denote the information in m

$$\imath_W(m) \overset{\triangle}{=} \log \frac{1}{P_W(m)}, \tag{17}$$

and note that by the memorylessness assumption in (18),

$$\imath_{X^n;Y^n}(a^n; b^n) = \sum_{i=1}^{n} \imath_{X;Y}(a_i; b_i). \tag{18}$$

In the achievability scheme of [3, Theorem 3], at time n the decoder observes Y^n, computes M information densities $\imath_{X^n;Y^n}(\mathsf{C}_1^n; Y^n), \ldots, \imath_{X^n;Y^n}(\mathsf{C}_M^n; Y^n)$, where $\mathsf{C}_1^n, \ldots, \mathsf{C}_M^n$ are the codewords, and stops once any of them exceeds a threshold. Instead of one common threshold, we assign lower thresholds to the more likely messages.

Code construction: We define the common randomness (revealed to the encoder and decoder before the transmission starts) to be a random variable U as follows:

$$P_U \triangleq \underbrace{P_{X^\infty} \times \ldots \times P_{X^\infty}}_{|\mathcal{W}|} \tag{19}$$

where X^∞ consists of i.i.d. copies drawn from (any) capacity-achieving input distribution. A realization of U defines $|\mathcal{W}|$ infinite dimensional vectors $\mathbf{C}_m^\infty \in \mathcal{A}^\infty$, $m \in \mathcal{W}$. Having observed that the value of W equals $m_0 \in \mathcal{W}$, the encoder transmits channel symbols X_n as follows:

$$X_n = \mathsf{f}_n(m_0) \overset{\triangle}{=} \mathbf{C}_{m_0}^n \tag{20}$$

At time n, the decoder computes the values

$$I_n(m) = \imath_{X^n;Y^n}(\mathsf{f}_n(m); Y^n) - \imath_W(m), \tag{21}$$

The decoder defines the stopping times:

$$\tau_m = \inf\{n \geq 0 : I_n(m) \geq \gamma\} \tag{22}$$

where $\gamma > 0$ is an arbitrary constant. The final decision $\widehat{W}$ is made by the decoder at the stopping time $\tau^\star$:

$$\tau^\star = \min_{m \in \mathcal{W}} \tau_m \tag{23}$$

$$\widehat{W} = g(Y^{\tau^\star}) = \operatorname*{argmin}_{m \in \mathcal{W}} \tau_m . \tag{24}$$

where the tie-breaking rule is immaterial.

Analysis: We claim that, averaged over U, we have:

$$\mathbb{P}[W \neq \widehat{W}] \leq \exp(-\gamma) \tag{25}$$

$$C \mathbb{E}[\tau^\star] \leq H(W) + \gamma + a_0 . \tag{26}$$

Abbreviate the stopping time of the true message as

$$\tau \stackrel{\triangle}{=} \tau_W . \tag{27}$$

The union bound results in, cf. [3, Theorem 3]:

$$\mathbb{P}[W \neq \widehat{W}|W = m_0] \leq \sum_{m \in \mathcal{W}\backslash\{m_0\}} \mathbb{P}[\tau_m \leq \tau|W = m_0] \tag{28}$$

$$\tag{29}$$

Due to memorylessness of the channel, $\imath_{X^n;Y^n}(X^n;Y^n) - nC$ is a martingale with jump size bounded by a_0 (defined in (15)). For each $j \in \mathbb{N}$, $\min\{\tau, j\}$ is a bounded stopping time, so by Doob's optional stopping theorem ([34, Theorem 10.10]) we have

$$C \mathbb{E}[\min\{\tau, j\}] - H(W) = \mathbb{E}\left[I_{\min\{\tau,j\}}(W)\right] \leq \gamma + a_0, \tag{30}$$

By the monotone convergence theorem, it follows from (30) that

$$C \mathbb{E}[\tau] - H(W) \leq \gamma + a_0. \tag{31}$$

Therefore, the stopping time τ has bounded expectation, and since the martingale $\imath_{X^n;Y^n}(X^n;Y^n) - nC$ also has bounded increments, Doob's optional stopping theorem applies to conclude

$$\mathbb{E}\left[\imath_{X^\tau;Y^\tau}(X^\tau;Y^\tau)\right] = C \mathbb{E}[\tau] . \tag{32}$$

Next, we have, by a change of measure argument, for every $m \neq m_0$:

$$\mathbb{P}[\tau_m = n|W = m_0] \leq \exp\{-\imath_W(m) - \gamma\}\mathbb{P}[\tau_m = n|W = m] . \tag{33}$$

Consequently, using (33), we have

$$\mathbb{P}[\tau_m \le \tau | W = m_0] \le \mathbb{P}[\tau_m < \infty | W = m_0] \tag{34}$$

$$= \sum_{n=0}^{\infty} \mathbb{P}[\tau_m = n | W = m_0] \tag{35}$$

$$\le \exp\{-\imath_W(m) - \gamma\}, \tag{36}$$

where we used that $\mathbb{P}[\tau < \infty] = \mathbb{P}[\tau_m < \infty | W = m] = 1$, which in turn follows from (31). Summing (36) over all $m \ne m_0$ and using (28) we get (25). Note that the reasoning in (33)–(36) generalizes that in [3, (111)-(118)] to nonequiprobable messages.

The estimate of average length in (26) follows from $\tau^\star \le \tau$ and (30). The result (16) follows by taking $\gamma = \log \frac{1}{\epsilon}$. ■

Remark 1. A slightly less sharp bound could also be derived via a *variable-length separated scheme*: compress W losslessly into a variable-length string $\{0,1\}^*$ with average length less than $H(W)$, cf. [35], then send the length via $O(\log H(W))$ channel symbols and very small probability of error and finally send the data bits.

Next, we extend the zero-error bound in [3, Theorem 11] to the case of non-equiprobable messages:

Theorem 3. *For every DMC with capacity C there exists a constant a_1 such that for every discrete random variable W there exists an $(\ell, 0)$ VLFT code with*

$$C\ell \le H(W) + a_1 \tag{37}$$

Proof: Without loss of generality, we assume that $H(W) < \infty$ and W takes values in positive integers. The codebook is a countable collection of infinite strings $\mathbf{C}_m^\infty$, $m = 1, 2, \ldots$. Given the codebook and the realization of $W = m_0$, the encoder sends

$$X_n = \mathsf{f}^n(m_0) \stackrel{\triangle}{=} \mathbf{C}_{m_0}^n \tag{38}$$

The decoder outputs m error-free at the stopping time $\tau^\star$ given by:

$$\tau^\star = \inf_n \left\{ I_n(m) > \max_{j \ne m} I_n(j) \right\}, \tag{39}$$

where

$$I_n(m) = \imath_{X^n;Y^n}(\mathbf{C}_m^n; Y^n) - \imath_W(m). \tag{40}$$

According to (39), if the true message is m the transmission stops at the first instant n when $I_n(m)$ exceeds all $I_n(j)$, $j \neq m$. Note that τ depends on the transmitted message m, as permitted by the paradigm of VLFT codes.

Analysis:

$$\mathbb{P}\left[\tau^\star > n | W = m\right] = \mathbb{P}\left[\bigcup_{j \neq m} \{I_n(j) > I_n(m)\} \,|\, W = m\right] \tag{41}$$

Applying the random coding argument, we now assume that the codebook strings $\mathbf{C}_1^\infty, \mathbf{C}_2^\infty, \ldots$ are drawn i.i.d. from $P_{X^\infty} = P_X \times P_X \times \ldots$, where P_X is the capacity-achieving channel input distribution. Denoting by $\bar{X}^n$ an independent copy of X^n and taking the expectation of the right side of (41) with respect to the codebook, we have

$$\mathbb{P}\left[\bigcup_{j \neq m} \left\{\frac{P_W(j)P_{Y|X}(Y|X_j)}{P_W(m)P_{Y|X}(Y|X_m)} \geq 1\right\} \,|\, W = m\right]$$

$$\leq \mathbb{E}\left[\min\left\{1, \sum_{j=1}^\infty \mathbb{P}\left[\frac{P_W(j)P_{Y^n|X^n}(Y|\bar{X}^n)}{P_W(m)P_{Y^n|X^n}(Y^n|X^n)} \geq 1 \,|\, X^n, Y^n\right]\right\}\right] \tag{42}$$

$$\leq \mathbb{E}\left[\min\left\{1, \sum_{j=1}^\infty \frac{P_W(j)}{P_W(m)} \frac{\mathbb{E}\left[P_{Y^n|X^n}(Y^n|\bar{X}^n)|Y^n\right]}{P_{Y^n|X^n}(Y^n|X^n)}\right\}\right] \tag{43}$$

$$= \mathbb{E}\left[\exp(-|\imath_{X^n;Y^n}(X^n;Y^n) - \imath_W(m)|^+)\right], \tag{44}$$

where we used the union bound and $\min\{1, a\} = \exp\left(-\left|\log \frac{1}{a}\right|^+\right)$. Applying (44) we get

$$\mathbb{E}\left[\tau^\star\right] = \sum_{n=0}^\infty \mathbb{P}\left[\tau^\star > n\right] \tag{45}$$

$$\leq \sum_{n=0}^\infty \mathbb{E}\left[\exp(-|\imath_{X^n;Y^n}(X^n;Y^n) - \imath_W(W)|^+)\right]. \tag{46}$$

Finally, (46) implies (37) by applying the result [3, (162)-(179)]:

$$\sum_{n=0}^\infty \mathbb{E}\left[\exp(-|\imath_{X^n;Y^n}(X^n;Y^n) - \gamma|^+)\right] \leq \frac{\gamma}{C} + a_1. \tag{47}$$

■

III. ASYMPTOTIC EXPANSIONS OF THE RATE-DISTORTION TRADEOFF

A. Definitions

We move from the setup of Section II where a discrete message is transmitted over the channel with feedback to a more general scenario, in which a, possibly analog, signal is transmitted over a channel with feedback, under a fidelity constraint. We will consider the following scenarios:

1) *excess distortion probability:* A VLF code transmitting memoryless source $S^k \in \mathcal{S}^k$ with reproduction alphabet $\hat{\mathcal{S}}^k$ and separable distortion measure $\mathsf{d} \colon \mathcal{S}^k \times \hat{\mathcal{S}}^k \mapsto [0, +\infty]$ is called a (k, ℓ, d, ϵ) excess-distortion code if the decoding time and the distortion satisfy

$$\mathbb{E}[\tau] \leq \ell \tag{48}$$

$$\mathbb{P}[\mathsf{d}(S^k, \hat{S}^k) > d] \leq \epsilon \tag{49}$$

The corresponding fundamental limit is

$$\ell^\star(k, d, \epsilon) \triangleq \inf\{\ell \colon \exists \text{ an } (k, \ell, d, \epsilon) \text{ VLF code}\}. \tag{50}$$

2) *average distortion:* A VLF code satisfying, instead of (49), an average constraint

$$\mathbb{E}[\mathsf{d}(S^k, \hat{S}^k)] \leq d \tag{51}$$

is called a (k, ℓ, d) average-distortion code. The corresponding fundamental limit is

$$\ell^\star(k, d) \triangleq \inf\{\ell \colon \exists \text{ an } (k, \ell, d) \text{ VLF code}\}. \tag{52}$$

3) *guaranteed distortion*: A VLFT code transmitting memoryless source $S^k \in \mathcal{S}^k$ with reproduction alphabet $\hat{\mathcal{S}}^k$ and separable distortion metric d is called a $(k, \ell, d, 0)$ guaranteed-distortion code if it achieves $\epsilon = 0$ in (49). The corresponding fundamental limit is

$$\ell_t^\star(k, d, 0) \triangleq \inf\{\ell \colon \exists \text{ an } (k, \ell, d, 0) \text{ VLFT code}\}. \tag{53}$$

We will use the following notation for the various minimizations of information measures:

$$\mathbb{R}_S(d) \triangleq \min_{\substack{P_{Z|S} \colon \mathcal{S} \mapsto \hat{\mathcal{S}} \colon \\ \mathbb{E}[\mathsf{d}(S,Z)] \leq d}} I(S; Z) \tag{54}$$

$$\mathbb{R}_S(d, \epsilon) \triangleq \min_{\substack{P_{Z|S} \colon \mathcal{S} \mapsto \hat{\mathcal{S}} \colon \\ \mathbb{P}[\mathsf{d}(S,Z) > d] \leq \epsilon}} I(S; Z) \tag{55}$$

$$H_{d,\epsilon}(S) \triangleq \min_{\substack{\mathsf{c} \colon \mathcal{S} \mapsto \hat{\mathcal{S}} \colon \\ \mathbb{P}[\mathsf{d}(S,\mathsf{c}(S)) > d] \leq \epsilon}} H(\mathsf{c}(S)). \tag{56}$$

The quantity in (56) is referred to as the (d, ϵ)-entropy of the source S [36]. The $(d, 0)$-entropy is also known as epsilon-entropy [36]:[6]

$$H_{d,0}(S) \triangleq \min_{\substack{\mathsf{c}: \mathcal{S} \mapsto \hat{\mathcal{S}}: \\ \mathsf{d}(S, \mathsf{c}(S)) \leq d \text{ a.s.}}} H(\mathsf{c}(S)). \tag{57}$$

Provided that the infimum in (54) is achieved by some transition probability kernel $P_{Z^\star|S}$, the d-tilted information in $s \in \mathcal{S}$ is defined as [24]

$$\jmath_S(s, d) \triangleq -\log \mathbb{E}\left[\exp\left(-\lambda^\star \mathsf{d}(s, Z^\star) + \lambda^\star d\right)\right] \tag{58}$$

where[7]

$$\lambda^\star = -\mathbb{R}'_S(d). \tag{59}$$

In almost-lossless compression, $Z^\star = S$ and

$$\jmath_S(s, d) \triangleq \imath_S(s). \tag{60}$$

B. Regularity assumptions on the source

We assume that the source, together with its distortion measure, satisfies the following assumptions:

A1 The source $\{S_i\}$ is stationary and memoryless, $P_{S^k} = P_\mathsf{S} \times \ldots \times P_\mathsf{S}$.

A2 The distortion measure is separable, $\mathsf{d}(s^k, z^k) = \frac{1}{k}\sum_{i=1}^{k} \mathsf{d}(s_i, z_i)$.

A3 The distortion level satisfies $d_{\min} < d < d_{\max}$, where $d_{\min}$ is the infimum of values at which the minimal mutual information quantity $\mathbb{R}_S(d)$ is finite, and $d_{\max} = \inf_{\mathsf{z} \in \hat{\mathcal{S}}} \mathbb{E}\left[\mathsf{d}(\mathsf{S}, \mathsf{z})\right]$, where the expectation is with respect to the unconditional distribution of S.

A4 The rate-distortion function is achieved by a unique $P_{\mathsf{Z}^\star|\mathsf{S}}$: $\mathbb{R}_\mathsf{S}(d) = I(\mathsf{S}; \mathsf{Z}^\star)$.

A5 $\mathbb{E}\left[\mathsf{d}^{12}(\mathsf{S}, \mathsf{Z}^\star)\right] < \infty$ where the expectation is with respect to $P_\mathsf{S} \times P_{\mathsf{Z}^\star}$.

The rate-dispersion function of the source satisfying assumptions A1–A5 is given by [24]

$$\mathcal{V}(d) = \mathrm{Var}\left(\jmath_\mathsf{S}(\mathsf{S}, d)\right). \tag{61}$$

We showed in [5] that under assumptions A1–A5 for all $0 \leq \epsilon \leq 1$

$$\left.\begin{array}{r}\mathbb{R}_{S^k}(d, \epsilon) \\[6pt] H_{d,\epsilon}(S^k)\end{array}\right\} = (1 - \epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}} e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O\left(\log k\right). \tag{62}$$

---

[6]N.B. in that terminology "epsilon" corresponds to d, not ϵ.

[7]Note that the existence of $P_{Z^\star|S}$ guarantees the differentiability of $\mathbb{R}_S(d)$.

C. Average distortion

Theorem 4. *Under assumptions A1–A5 we have*

$$C\ell^*(k, d) = kR(d) + O(\log k). \tag{63}$$

Proof:

Achievability: fix $1 < p \leq 12$, $\epsilon > 0$, source codebook $c_1, \ldots, c_M$ and consider a separated scheme $S - \mathsf{f}(S) - X - Y - \widehat{W} - \mathsf{c}(\widehat{W})$ such that $\mathbb{P}\left[\mathsf{f}(S) \neq \widehat{W}\right] \leq \epsilon$ and

$$\mathsf{f}(s^k) = \arg \min_{m \in \{1, \ldots, M\}} \mathsf{d}(s^k, c_m), \quad s^k \in \mathcal{S}^k \tag{64}$$

$$\mathsf{c}(m) = c_m, \quad m \in \{1, \ldots, M\} \tag{65}$$

The average distortion is bounded by

$$\mathbb{E}\left[\mathsf{d}\left(S^k, c_{\widehat{W}}\right)\right] \leq \mathbb{E}\left[\mathsf{d}(S^k, c_W)\right] + \mathbb{E}\left[\mathsf{d}(S^k, c_{\widehat{W}})1\left\{\widehat{W} \neq W\right\}\right] \tag{66}$$

$$\leq \mathbb{E}\left[\min_{m \in \{1, \ldots, M\}} \mathsf{d}(S^k, c_m)\right] + \left(\mathbb{E}\left[\mathsf{d}^p(S^k, c_{\widehat{W}})\right]\right)^{\frac{1}{p}} \epsilon^{1 - \frac{1}{p}} \tag{67}$$

where (67) holds by Hölder's inequality. Taking the expectation of both sides over $c_1, \ldots, c_M$ drawn i.i.d. from $P_{\mathsf{Z}^\star}^k$, we conclude, via a random coding argument, that there exists a separate source/channel code with average distortion bounded by

$$d \leq \bar{d} + \left(\mathbb{E}\left[\mathsf{d}^p(S^k, Z_1)\right]\right)^{\frac{1}{p}} \epsilon^{1 - \frac{1}{p}}, \tag{68}$$

where we denoted

$$\bar{d} \triangleq \mathbb{E}\left[\min_{m \in \{1, \ldots, M\}} \mathsf{d}(S^k, Z_m)\right], \tag{69}$$

and $P_{S^k Z_1 \ldots Z_M} = P_{S^k} P_{\mathsf{Z}^\star}^k \ldots P_{\mathsf{Z}^\star}^k$. The work of Pilc [37] (finite alphabet) and Yang and Zhang [38] (abstract alphabet) shows that under assumptions A1–A5,[8]

$$\log M = kR(\bar{d}) + O(\log k). \tag{70}$$

Letting $\epsilon = k^{\frac{p}{p-1}}$, we conclude by assumption A5 that the second term in (68) is bounded by $O\left(\frac{1}{k}\right)$, i.e. $d \leq \bar{d} + O\left(\frac{1}{k}\right)$. Finally, by Theorem 1, $C\ell \leq \log M + O(\log k)$, and the '$\leq$' direction in (63) follows by the differentiability of $R(d)$ in the region of assumption A3.

---

[8]In particular, the bounded twelfth moment in A5 is required for the applicability of the result of Yang and Zhang [38].

Converse: By the data-processing inequality and [2, Lemma 1-2] we have

$$kR(d) \leq \ell C \tag{71}$$

for any (k, ℓ, d) VLF code. ∎

D. Excess distortion probability

Theorem 5. *Under assumptions A1–A5 and any $\epsilon > 0$ we have*

$$\ell^\star(k, d, \epsilon)\, C = (1 - \epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}} e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O\left(\log k\right) \tag{72}$$

Proof: Achievability: Pair a lossy compressor $S^k \to W$ with excess-distortion probability $\epsilon' = \epsilon - \frac{1}{\sqrt{k}}$ and $H(W) = H_{d,\epsilon'}(S^k)$[9] with a VLF code from Theorem 2 transmitting W with probability of error $\frac{1}{\sqrt{k}}$. Apply (62) to (16)[10].

Converse: Apply the data-processing inequality and [2, Lemma 1-2] to get:

$$\ell C \geq \mathbb{R}_{S^k}(d, \epsilon) \tag{73}$$

for every (k, ℓ, d, ϵ) VLF code. ∎

E. Guaranteed distortion

Theorem 6. *Under assumptions A1–A5, we have*

$$\ell_t^\star(k, d, 0)C = kR(d) + O\left(\log k\right) \tag{74}$$

Proof: For the achievability we note that the estimate of the $H_{d,\epsilon}(S^k)$ in (62) applies with $\epsilon = 0$ and thus

$$H_{d,0}(S^k) = kR(d) + O(\log k). \tag{75}$$

---

[9]Although the optimal mapping $c^\star$ that achieves (d, ϵ)-entropy is not known in general, the existence of good approximations satisfying the constraint in (56) that approach $H_{d,\epsilon}(S^k)$ to within $\log_2 H_{d,\epsilon}(S^k)$ bits is guaranteed by a random coding argument, see [5].

[10]Note that (62) also holds if ϵ in the left side is replaced by $\epsilon + O\left(\frac{1}{\sqrt{k}}\right)$.

Then, we can pair the mapping achieving $H_{d,0}(S^k)$ with the zero-error VLFT code from Theorem 3.

Conversely, repeating the argument of [3, Theorem 4], with the replacement of the right side of [3, (67)] by $\mathbb{R}_S(d, \epsilon)$ we conclude that any (ℓ, d, ϵ) VLFT code must satisfy

$$\mathbb{R}_S(d, \epsilon) \leq C\ell + \log(\ell + 1) + \log e. \tag{76}$$

∎

F. Discussion

We make several remarks regarding the rate-distortion tradeoff in all three settings considered above:

1) The case $d = d_{\min}$ is excluded by the assumptions of Theorems 4–6. However, in the important special case of a distortion measure that satisfies

$$\mathsf{d}(a, b) = \begin{cases} d_{\min}, & a = b \\ > d_{\min}, & a \neq b \end{cases}, \tag{77}$$

$d = d_{\min}$ corresponds to almost-lossless transmission, and both Theorems 4 and 5 apply with $R(d)$ and $\mathcal{V}(d)$ equal to the entropy and the varentropy of the source, respectively, as long as the source is stationary and memoryless and the third moment of $\imath_S(\mathsf{S})$ is finite.

2) For almost-lossless transmission of finite alphabet sources, the asymptotic expansion (72) can be achieved by reliably (i.e. with probability of error $\sim \frac{1}{k}$) sending through the channel the type of the source outcome first, and then reliably sending each message whose type is one the most likely types with total mass $1 - \epsilon$.

3) Even if the channel is not symmetric, the asymptotic expansions in Theorems 4–6 can be achieved without common randomness U, by using constant composition channel codebooks. For instance, consider the scheme in Theorem 2 with P_{X^∞} drawn from the distribution equiprobable over the capacity-achieving type. Since $\mathbb{E}[\tau|X^\infty] = \mathbb{E}[\tau]$ a.s., almost every such codebook attains the bound in (30) up to logarithmic terms, resulting in a deterministic construction attaining (72).

4) Stop-feedback codes are remarkably powerful at finite blocklength; indeed, up to the terms of order $O(\log k)$, they attain the fundamental limits in the settings of Theorems 4, 5 and

6. As the converse parts of Theorems 4, 5 and 6 demonstrate, relying on feedback more heavily can only bring in an improvement of order at most $O(\log k)$.

5) Note that (72) is achieved by a stop-feedback code. We can further show that even without any feedback one can still achieve the optimal first order performance

$$\ell C \leq (1 - \epsilon)kR(d) + O(\sqrt{k \log k}), \tag{78}$$

provided variable-length channel coding is allowed. Indeed, one can first use the variable-length excess-distortion compressor from [5] on S^k to get a binary string of average length $(1-\epsilon)kR(d)+O(\sqrt{k})$, see (2). Then, truncating the length at k^2 and transmitting $2\log k$ data bits with reliability $\frac{1}{k^2}$, we can reliably inform the encoder about the total number of data bits b to be sent next. We may then use a capacity-achieving code of length $\frac{b}{C}+O(\sqrt{b \log b})$ to send the data bits with reliability $\frac{1}{k}$ [23].

6) The naive separation scheme, i.e. a fixed-length source code followed by a channel code achieves at most:

$$\ell C \geq (1 - \epsilon)kR(d) + a\sqrt{k \log k}, \, a > 0. \tag{79}$$

Indeed, according to Theorem 1, the number of messages M that can be transmitted via a VLF code with error probability η satisfies

$$\log M \geq \frac{\ell C}{1 - \eta} + O(\log \ell). \tag{80}$$

On the other hand, the number of codewords of a source code with probability of exceeding distortion d no greater than ζ satisfies [24]

$$\log M \leq kR(d) + \sqrt{k\mathcal{V}(d)}Q^{-1}(\zeta) + O(1). \tag{81}$$

Optimizing over $\eta + \zeta \leq \epsilon$ yields (79).

7) The semi-joint separated schemes that attain (72) contain a vital ingredient missing from naive separated schemes: namely, the channel code employs unequal error protection. Consequently, the more likely source codewords are decoded with higher reliability, resulting in massive improvement at finite blocklength evidenced by (72). Unequal error protection can be achieved ether via a maximum-a-posteriori-like decoder of Theorem 2 or the variable-length separation interface of Remark 1.

8) The Schalkwijk-Bluestein [22] (see also [39]) elegant linear feedback scheme for the transmission of a single Gaussian sample $S \sim \mathcal{N}(0, \sigma^2)$ over the AWGN channel achieves the

mean-square error $\frac{\sigma^2}{(1+P)^n}$, after n channel uses, where P is the average transmit SNR. In other words, the minimum delay in transmitting a Gaussian sample over a Gaussian channel with feedback is given by

$$\ell^\star(1, d) = \frac{R(d)}{C}, \tag{82}$$

as long as $\frac{R(d)}{C}$ is integer.[11] Note that (82) is achieved with fixed, not variable length, and average, not maximal, power constraint. If there are k Gaussian samples to transmit, repeating the scheme for each of the samples achieves

$$\ell^\star(k, d) = k\frac{R(d)}{C}, \tag{83}$$

which implies, in particular, that in general our estimate of $O\left(\log k\right)$ in (63) is too conservative. Beyond Gaussian sources and channels, a sufficient condition for a fixed-length JSCC feedback scheme to achieve (83) is provided in [16].

9) The Schalkwijk-Bluestein scheme uses instantaneous feedback and has notoriously resisted generalization beyond Gaussian channels, which limits the applicability of the scheme. In contrast, the simple separated scheme in Theorem 4 uses only stop-feedback and applies to arbitrary sources and channels.

IV. ENERGY-LIMITED FEEDBACK CODES FOR NON-EQUIPROBABLE MESSAGES

In this section, we study the transmission of a message over an AWGN channel under an energy constraint. We would like to know how much information can be pushed through the channel, if a total of E units of energy is available to accomplish the task. Formally, the codes studied in this section are defined as follows.

Definition 3. *An energy-limited code without feedback for the transmission of a random variable* W *taking values in* $\mathcal{W}$ *over an AWGN channel is defined by:*

1) A sequence of encoders $\mathsf{f}_n \colon \mathcal{W} \mapsto \mathcal{A}$, *specifying the channel inputs*

$$X_n = \mathsf{f}_n\left(W\right) \tag{84}$$

satisfying

$$\mathbb{P}\left[\sum_{j=1}^{\infty} X_j^2 \leq E\right] = 1 \tag{85}$$

---

[11]If $R(d) = C$, no feedback is needed.

2) A decoder g: $\mathcal{B}^\infty \mapsto \mathcal{W}$.

Definition 4. *An energy-limited feedback code for the transmission of a random variable W taking values in $\mathcal{W}$ over an AWGN channel is defined by:*

1) A sequence of encoders $\mathsf{f}_n\colon \mathcal{W} \times \mathcal{B}^{n-1} \mapsto \mathcal{A}$, *defining the channel inputs*

$$X_n = \mathsf{f}_n\left(W, Y^{n-1}\right) \tag{86}$$

satisfying

$$\sum_{j=1}^{\infty} \mathbb{E}\left[X_j^2\right] \leq E \tag{87}$$

2) A decoder g: $\mathcal{B}^\infty \mapsto \mathcal{W}$.

An (E, ϵ) code for the transmission of random variable W over the Gaussian channel is a code with energy bounded by E and $\mathbb{P}\left[W \neq \widehat{W}\right] \leq \epsilon$.

Definitions 3–4 do not impose any restrictions on the number of degrees of freedom n, restricting instead the total available energy. The problem of transmitting a message with minimum energy was posed by Shannon [40], who showed that E_b, the minimum energy per information bit compatible with vanishing block error probability converges to $N_0 \log_e 2$ as the number of information bits goes to infinity, where $\frac{N_0}{2}$ is the noise power per degree of freedom. Recently, Polyanskiy et al. [30, Theorem 7] showed a dynamic programming algorithm for the error-free transmission of a single bit over an AWGN channel with feedback that attains *exactly* Shannon's optimal energy per bit tradeoff

$$E_b = N_0 \log_e 2. \tag{88}$$

The next non-asymptotic achievability result leverages that algorithm to transmit error-free a binary representation of a random variable over the AWGN channel by means of a variable-length separate compression/transmission scheme.

Theorem 7. *There exists a zero-error feedback code for the transmission of a random variable W over the AWGN channel with energy*

$$\frac{E}{N_0} \log e < H(W) + 1. \tag{89}$$

Conversely, any $(E, 0)$-feedback code must satisfy

$$H(W) \leq \frac{E}{N_0} \log e. \tag{90}$$

Bit number	Sequence of time slots
1	1 2 4 7 $\cdots$
2	3 5 8 $\cdots$
3	6 9 $\cdots$
$\vdots$	
$\ell(W)$	

Fig. 1. Illustration of the diagonal numbering of channel uses in Theorem 7.

Proof: The encoder converts the source into a variable-length string using the Huffman code, so that the codebook is prefix-free and the expectation of the encoded length $\ell(W)$ is bounded as

$$\mathbb{E}\left[\ell(W)\right] < H(W) + 1. \tag{91}$$

Next each bit (out of $\ell(W)$) is transmitted at the optimal energy per bit tradeoff $N_0 \log_e 2$ using the zero-error feedback scheme in [30, Theorem 7]. Transmissions corresponding to different bits are interleaved diagonally (see Fig. 1): the first bit is transmitted in time slots $1, 2, 4, 7, 11, \ldots$, the second one in $3, 5, 8, 12, \ldots$, and so on. The channel encoder is silent at those indices allocated to source bits $\ell(W) + 1, \ell(W) + 2, \ldots$ For example, if the codeword has length 2 nothing is transmitted in time slots $6, 9, 13, \ldots$. The receiver decodes the first transmitted bit focusing on the time slots $1, 2, 4, 7, 11, \ldots$ It proceeds successively with the second bit, etc., until it forms a codeword of the Huffman code, at which point it halts. Thus, it does not need to examine the outputs of the time slots corresponding to information bits that were not transmitted, and in which the encoder was silent.

Since the scheme spends $N_0 \log_e 2$ energy per bit, the total energy to transmit the codeword representing W is

$$\ell(W) N_0 \log_e 2. \tag{92}$$

Taking the expectation of (92) over W and applying (91), (89) follows.

In the converse direction, due to the zero-error requirement and data processing, $H(W) = I(W; \mathsf{g}(Y^\infty)) \leq I(X^\infty; Y^\infty)$, which in turn is bounded by the right side of (90) [40]. ∎

Our next achievability result studies the performance of a variable-length separated scheme.

Theorem 8. *Fix positive E_1 and E_2 such that*

$$E_1 + E_2 \leq E. \tag{93}$$

Denote

$$\varepsilon(E, m) \triangleq 1 - \frac{1}{\sqrt{\pi N_0}} \int_{-\infty}^{\infty} \left(1 - Q\left(\frac{x + \sqrt{E}}{\sqrt{\frac{N_0}{2}}}\right)\right)^{m-1} e^{-\frac{x^2}{N_0}} \, dx. \tag{94}$$

Assume that W takes values in $\{1, 2, \ldots, M\}$. There exists an (E, ϵ) non-feedback code for the transmission of random variable W over an AWGN channel without feedback such that

$$\epsilon \leq \mathbb{E}\left[\varepsilon\left(E_1, W\right)\right] + \varepsilon(E_2, \lfloor \log_2 M \rfloor + 1). \tag{95}$$

Proof: Assume that the outcomes of W are ordered in decreasing probabilities. Consider the following variable-length separated achievability scheme: the source outcome m is first losslessly represented as a binary string of length $\lfloor \log_2 m \rfloor$ by assigning it to m-th binary string in $\{\varnothing, 0, 1, 00, 01, \ldots\}$ (the most likely outcome is represented by the empty string). Then, all binary strings are grouped according to their encoded lengths. A channel codebook is generated for each group of sequences. The encoded length is sent over the channel with high reliability, so the decoder almost never makes an error in determining that length. Then the encoder makes an ML decision only between sequences of that length. A formal description and an error analysis follow.

Codebook: the collection of $M + \lfloor \log M \rfloor + 1$ codewords

$$\mathbf{c}_j = \sqrt{E_1} \, \mathbf{e}_j, \ \ j = 1, 2, \ldots, M \tag{96}$$

$$\mathbf{c}_j = \sqrt{E_2} \, \mathbf{e}_j, \ \ j = M + 1, \ldots, M + \lfloor \log_2 M \rfloor + 1 \tag{97}$$

where $\{\mathbf{e}_j, \ j = 1, 2, \ldots\}$ is an orthonormal basis of $L_2(\mathbb{R}^{\infty})$.

Encoder: The encoder sends the pair $(m, \lfloor \log_2 m \rfloor)$ by transmitting $\mathbf{c}_m + \mathbf{c}_{M + \lfloor \log_2 m \rfloor + 1}$.

Decoder: Having received the infinite string corrupted by i.i.d. Gaussian noise $\mathbf{z}$, the decoder first (reliably) decides between $\lfloor \log_2 M \rfloor + 1$ possible values of $\lfloor \log_2 m \rfloor$ based on the minimum distance:

$$\hat{\ell} \triangleq \underset{j}{\operatorname{argmin}} \, \|\mathbf{z} - \mathbf{c}_{M+j+1}\|, \ \ j = 0, \ldots, \lfloor \log_2 M \rfloor \tag{98}$$

As shown in [41, p. 258]), [42], [30, Theorem 3], the probability of error of such a decision is given by $\varepsilon(E, \lfloor \log_2 M \rfloor + 1)$. This accounts for the second term in (95). The decoder then decides between $2^{\hat{\ell}}$ messages[12] j with $\lfloor \log j \rfloor = \hat{\ell}$:

$$\hat{\mathbf{c}} \triangleq \arg\min \|\mathbf{z} - \mathbf{c}_j\|, \ j = 2^{\hat{\ell}}, \ldots, \min\{2^{\hat{\ell}+1} - 1, M\} \tag{99}$$

The probability of error of this decision rule is similarly upper bounded by $\varepsilon(E, m)$, provided that the value of $\lfloor \log_2 m \rfloor$ was decoded correctly: $\hat{\ell} = \lfloor \log_2 m \rfloor$. Since $2^{\lfloor \log_2 m \rfloor} \leq m$, this accounts for the first term in (95).

$\blacksquare$

Normally, one would choose $1 \ll E_2 \ll E_1$ so that the second term in (95), which corresponds to the probability of decoding the length incorrectly, is negligible compared to the first term, and the total energy $E \approx E_1$. Moreover, if W takes values in a countably infinite alphabet, one can truncate it so that the tail is negligible with respect to the first term in (95). To ease the evaluation of the first term in (95), one might use $i \leq \frac{1}{P_W(i)}$. In the equiprobable case, this weakening leads to $\mathbb{E}\left[\varepsilon(E_1, W)\right] \leq \varepsilon(E_1, M)$.

If the power constraint is average rather than maximal, a straightforward extension of Theorem 8 ensures the existence of an (E, ϵ) code (average power constraint) for the AWGN channel with

$$\epsilon \leq \mathbb{E}\left[\varepsilon\left(E_1(\lfloor \log_2 W \rfloor), W\right)\right] + \varepsilon(E_2, \lfloor \log M \rfloor + 1), \tag{100}$$

where $E_1 \colon \{0, 1, \ldots, \lfloor \log_2 M \rfloor\} \mapsto \mathbb{R}_+$ and $E_2 \in \mathbb{R}_+$ are such that

$$\mathbb{E}\left[E_1(\lfloor \log_2 W \rfloor)\right] + E_2 \leq E. \tag{101}$$

V. ASYMPTOTIC EXPANSIONS OF THE ENERGY-DISTORTION TRADEOFF

A. Problem setup

This section focuses on the energy-distortion tradeoff in the JSCC problem. Like in Section IV, we limit the total available transmitter energy E without any restriction on the (average) number of channel uses per source sample. Unlike Section IV, we allow general (not neccesarily discrete) sources, and we study the tradeoff between the source dimension k, the total energy E and the fidelity of reproduction. Thus, we identify the minimum energy compatible with target

---

[12]More precisely, $2^{\hat{\ell}}$ messages if $\hat{\ell} \leq \lfloor \log_2 M \rfloor - 1$ and $M - 2^{\lfloor \log_2 M \rfloor} + 1 \leq 2^{\lfloor \log_2 M \rfloor}$ messages if $\hat{\ell} = \lfloor \log_2 M \rfloor$.

distortion without any restriction on the time-bandwidth product (number of degrees of freedom). As in Section III-A, we consider both the average and the excess distortion criteria.

Formally, we let the source be a k-dimensional vector $S^k \in \mathcal{S}^k$. A (k, E, d, ϵ) energy-limited code is an energy-limited code for S^k with total energy E and probability $\leq \epsilon$ of distortion exceeding d (see (49)). Similarly, a (k, E, d) energy-limited code is an energy-limited code for S^k with total energy E and average distortion not exceeding d (see (51)). The goal of Section V is to characterize the minimum energy required to transmit k source samples at a given fidelity, i.e. to characterize the following fundamental limits:

$$E_f^\star(k, d) \triangleq \{\inf E \colon \exists \text{ a } (k, E, d) \text{ feedback code}\}, \tag{102}$$

$$E_f^\star(k, d, \epsilon) \triangleq \{\inf E \colon \exists \text{ a } (k, E, d, \epsilon) \text{ feedback code}\} \tag{103}$$

as well as the corresponding limits $E^\star(k, d)$ and $E^\star(k, d, \epsilon)$ of the energy-limited non-feedback codes.

B. Previous results on the energy-per-bit and the energy-distortion tradeoff

If the source produces equiprobable binary strings of length k, Shannon [40] showed that the minimum energy per information bit to noise power spectral density ratio compatible with vanishing block error probability converges to

$$\frac{E^\star(k, 0, \epsilon)}{kN_0} \to \log_e 2 = -1.59 \text{ dB} \tag{104}$$

as $k \to \infty$, $\epsilon \to 0$. The fundamental limit in (104) holds regardless of whether feedback is available. Moreover, this fundamental limit is known to be the same regardless of whether the channel is subject to fading or whether the receiver is coherent or not [43]. Polyanskiy et al. refined (104) as [30, Theorem 3]

$$E^\star(k, 0, \epsilon)\frac{\log e}{N_0} = k + \sqrt{2k \log e}\, Q^{-1}(\epsilon) - \frac{1}{2}\log k + O\left(1\right) \tag{105}$$

for transmission without feedback, and as [30, Theorem 8]

$$E_f^\star(k, 0, \epsilon)\frac{\log e}{N_0} = (1 - \epsilon)k + O\left(1\right) \tag{106}$$

for transmission with feedback. Moreover, [30, Theorem 7] (see also (88)) shows that in fact

$$E^\star(k, 0, 0)\frac{\log e}{N_0} = k, \tag{107}$$

i.e. in the presence of full noiseless feedback, Shannon's limit (104) can be achieved with equality already at $k = 1$ and $\epsilon = 0$.

For the finite blocklength behavior of energy per bit in fading channels, see [44].

For the transmission of a memoryless source over the AWGN channel under an average distortion criterion, Jain et al. [29, Theorem 1] pointed out that as $k \to \infty$,

$$\frac{E^\star(k, d)}{k} \frac{\log e}{N_0} \to R(d). \tag{108}$$

Note that (108) still holds even if noiseless feedback is available.

Unlike Polyanskiy et al. [30], we allow analog sources and arbitrary distortion criteria, and unlike Jain et al. [29], we are interested in a nonasymptotic analysis of the minimum energy per sample.

C. Energy-limited feedback codes

Our first result in this section is a refinement of (108).

Theorem 9. *Let the source and its distortion measure satisfy assumptions A1–A5. The minimum energy required to transmit k source symbols with average distortion $\leq d$ over an AWGN channel with feedback satisfies*

$$E_f^\star(k, d) \cdot \frac{\log e}{N_0} = kR(d) + O(\log k) \tag{109}$$

Proof: Achievability. The expansion in (109) is achieved by the following separated source/channel scheme. For the source code, we use the code of Yang and Zhang [38] (abstract alphabet) that compresses the source down to M representation points with average distortion d such that

$$\log M = kR(d) + O(\log k). \tag{110}$$

For the channel code, we transmit the binary representation of M error-free using the optimal scheme of Polyanskiy et al. [30, Theorem 7], so that

$$\log M = \frac{E}{N_0} \log e. \tag{111}$$

Converse. By data processing, similar to (90),

$$kR(d) \leq \frac{E}{N_0} \log e. \tag{112}$$

∎

Remark 2. For the transmission of a Gaussian source over the feedback AWGN channel, we have

$$E_f^\star(k, d) \cdot \frac{\log e}{N_0} = kR(d). \tag{113}$$

Indeed, the Schalkwijk-Bluestein scheme [22], [39] attains (113) for $k = 1$. For $k > 1$, transmitting the Schalkwijk-Bluestein codewords corresponding to i-th source sample in time slots $i, k + i, 2k + i, \ldots$ attains (113) exactly for all $k = 1, 2, \ldots$.

Theorem 10. *In the transmission of a source satisfying the assumptions A1–A5 over an AWGN channel with feedback, the minimum average energy required for the transmission of k source samples under the requirement that the probability of exceeding distortion d is no greater than $0 \le \epsilon < 1$ satisfies, as $k \to \infty$,*

$$E_f^\star(k, d, \epsilon) \frac{\log e}{N_0} = (1 - \epsilon)kR(d) - \sqrt{\frac{k\mathcal{V}(d)}{2\pi}} e^{-\frac{(Q^{-1}(\epsilon))^2}{2}} + O(\log k) \tag{114}$$

Proof: Achievability. Pair a lossy compressor $S^k \to W$ with excess-distortion probability ϵ and $H(W) = H_{d,\epsilon}(S^k)$ with the achievability scheme in Theorem 7 and apply (89) and (62).

Converse. Again, the converse result follows proceeding as in (90), invoking (62). ■

Comparing (114) and (72), we observe that, similar to the setup in Section III, allowing feedback and average power constraint reduces the asymptotically achievable minimum energy per sample by a factor of $1 - \epsilon$. As in Section III, that limit is approached from below rather than from above, i.e. finite blocklength helps.

Similar to the setup of Section III, naive separation achieves at most

$$E_f^\star(k, d, \epsilon) \frac{\log e}{N_0} \le (1 - \epsilon)kR(d) + a\sqrt{k \log k}, \ a > 0. \tag{115}$$

D. Energy-limited non-feedback codes

Our next result generalizes [30, Theorem 3]. Loosely speaking, it shows that the energy E, probability of error ϵ and distortion d of the best non-feedback code satisfy

$$\frac{E}{N_0} \log e - kR(d) \approx \sqrt{k\mathcal{V}(d) + \frac{2E}{N_0} \log e} \cdot Q^{-1}(\epsilon). \tag{116}$$

Note that in (116) source and channel dispersions add up, as in the usual (non-feedback) joint source-channel coding problem [6], [8]. More precisely, we have the following:

Theorem 11. *In the transmission of a stationary memoryless source (satisfying the assumptions A1–A5) over the AWGN channel, the minimum energy necessary for achieving probability $0 < \epsilon < 1$ of exceeding distortion d satisfies, as $k \to \infty$,*

$$E^\star(k, d, \epsilon) \frac{\log e}{N_0} = kR(d) + \sqrt{k(2R(d)\log e + \mathcal{V}(d))}Q^{-1}(\epsilon) + O(\log k) \qquad (117)$$

Proof:

Achievability: We let the total energy E be such that

$$E\frac{\log e}{N_0} = kR(d) + \sqrt{k(2R(d)\log e + \mathcal{V}(d))}Q^{-1}\left(\epsilon - \frac{a}{\sqrt{k}}\right) + b\log k, \qquad (118)$$

and we show that $a > 0$ and b can be chosen so that the excess distortion probability is bounded by ϵ.

We consider a good lossy code with $M = \exp(2kR(d))$ representation points, so that the probability that the source is not represented within distortion d is exponentially small. We show that a combination of that code with the variable-length separated scheme in Theorem 8 achieves (118). First, we prove the following generalization of Theorem 8 to the lossy case: for any M, there exists an (k, E, d, ϵ') code for the AWGN channel (without feedback) such that

$$\epsilon' \leq \mathbb{E}\left[\varepsilon\left(E_1, \frac{1}{P_{\mathsf{Z}^\star}^k(B_d(S^k))}\right)\right] + \varepsilon(E_2, \lfloor \log M \rfloor + 1) + \mathbb{E}\left[\left(1 - P_{\mathsf{Z}^\star}^k(B_d(S^k))\right)^M\right], \quad (119)$$

where

$$B_d(s^k) \triangleq \{z^k \in \hat{\mathcal{S}}^k \colon \mathsf{d}(s^k, z^k) \leq d\}. \qquad (120)$$

Towards that end, let the representation points $Z_1, Z_2, \ldots, Z_M$ be drawn i.i.d. from $P_{\mathsf{Z}^\star}^k$. The source encoder goes down the list of the representation points and outputs the index of the first d-close match to S^k:

$$W \triangleq \min\{i \colon \mathsf{d}(S^k, Z_i) \leq d\} \qquad (121)$$

(if there is no such index, it outputs 1). Averaged over $Z_1, \ldots, Z_M$, the probability that no d-close match is found is upper bounded by the third term in (119) (e.g. [24, Theorem 10]). The index W is then transmitted over the channel using the scheme in Theorem 8, with the total probability of error averaged over all lossy codebooks given by

$$\epsilon' \leq \mathbb{E}\left[\varepsilon(E_1, W)\right] + \varepsilon(E_2, \lfloor \log M \rfloor + 1) + \mathbb{E}\left[\left(1 - P_{\mathsf{Z}^\star}^k(B_d(S^k))\right)^M\right] \qquad (122)$$

Since conditioned on $S = s$, W is geometrically distributed with success probability $P_{Z^\star}^k(B_d(s^k))$, we have

$$\mathbb{E}\left[W|S^k = s^k\right] = \frac{1}{P_{Z^\star}^k(B_d(s^k))} \tag{123}$$

Noting that $\varepsilon(E, m)$ is a concave function of m, we have by Jensen's inequality

$$\mathbb{E}\left[\varepsilon(E, W)\right] \leq \mathbb{E}\left[\varepsilon\left(E, \mathbb{E}\left[W|S^k\right]\right)\}\right], \tag{124}$$

which gives the first term in (95).

We proceed to show that with the choice of

$$E_1 = E - c\log E, \tag{125}$$

for an appropriate $c > 0$, and $M = \exp(2kR(d))$, the right side of (119) is upper bounded by ϵ.

A reasoning similar to [24, (108)–(111)] and the Cramér-Chernoff bound yield

$$\mathbb{E}\left[\left(1 - P_{Z^{k\star}}(B_d(S^k))\right)^M\right] \leq \exp(-ka_1) \tag{126}$$

for some $a_1 > 0$. On the other hand, (105) [30, Theorem 3] implies

$$\varepsilon(E, m) = \mathbb{P}\left[\log m > G(E)\right] + O\left(\frac{1}{\sqrt{k}}\right). \tag{127}$$

where $G(E) \overset{D}{=} \mathcal{N}\left(\frac{E}{N_0}\log e - \frac{1}{2}\log\frac{E}{N_0}, \frac{2E}{N_0}\log^2 e\right)$

Applying (126) and (127) to (95), we conclude that the excess-distortion probability is bounded above by

$$\mathbb{P}\left[\log\frac{1}{P_{Z^{k\star}}(B_d(S^k))} > G(E - c\log E)\right] + \mathbb{P}\left[\log(\log M + 1) > G(c\log E)\right] + O\left(\frac{1}{\sqrt{E}}\right). \tag{128}$$

The second term in (128) can be made to decay as fast as $O\left(\frac{1}{\sqrt{E}}\right)$ for large enough c. To evaluate the first term in (128), we recall [24, Lemma 2], which states that for k large enough,

$$\mathbb{P}\left[\log\frac{1}{P_{Z^{k\star}}(B_d(S^k))} \leq \sum_{i=1}^{k} \jmath_{\mathsf{S}}(S_i, d) + C\log k + c_0\right] \geq 1 - \frac{K}{\sqrt{k}} \tag{129}$$

where c_0 and K are constants, and C is a constant explicitly identified in [24, Lemma 2]. Letting $b = c + C - \frac{1}{2}$ and applying (129) to upper-bound the first term in (128), we conclude by the Berry-Esséen Theorem that

$$\mathbb{P}\left[\log\frac{1}{P_{Z^{k\star}}(B_d(S^k))} > G(E - c\log E)\right] \leq \epsilon - \frac{a}{\sqrt{k}} + O\left(\frac{1}{\sqrt{k}}\right), \tag{130}$$

Since a can be chosen so that (128) is upper bounded by ϵ for k large enough, this concludes the proof of the achievability part.

Converse: The result in [6, Theorem 1] states that:

$$\epsilon \geq \sup_{\gamma > 0} \left\{ \sup_{P_{\bar{\mathbf{Y}}}} \mathbb{E} \left[\inf_{\mathbf{x}: \|\mathbf{x}\|^2 \leq E} \mathbb{P} \left[\jmath_{S^k}(S^k, d) - \imath_{\mathbf{X}; \bar{\mathbf{Y}}}(\mathbf{x}; \mathbf{Y}) \geq \gamma \mid S \right] \right] - \exp(-\gamma) \right\} \tag{131}$$

where $\jmath_{S^k}(s^k, d)$ is the d-tilted information in $s^k \in \mathcal{S}^k$ defined in (58), $\imath_{\mathbf{X}; \bar{\mathbf{Y}}}(\mathbf{x}; \mathbf{y}) \triangleq \log \frac{dP_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}}{dP_{\bar{\mathbf{Y}}}}(\mathbf{y})$, and $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ and $P_{\bar{\mathbf{Y}}}$ are specialized to

$$P_{\bar{\mathbf{Y}}} \overset{D}{=} \prod_{k=1}^{\infty} \mathcal{N}\left(0, \frac{N_0}{2}\right) \tag{132}$$

$$P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}} \overset{D}{=} \prod_{k=1}^{\infty} \mathcal{N}\left(x_k, \frac{N_0}{2}\right) \tag{133}$$

$$\imath_{\mathbf{X}; \bar{\mathbf{Y}}}(\mathbf{x}; \mathbf{Y}) \overset{D}{=} \mathcal{N}\left(\frac{\|\mathbf{x}\|^2}{N_0} \log e, \frac{2\|\mathbf{x}\|^2}{N_0} \log e\right) \tag{134}$$

Next, we let in (131) $\gamma = \frac{1}{2} \log \frac{E}{N_0}$. Since $\jmath_{S^k}(S^k, d) = \sum_{i=1}^{k} \jmath_{\mathsf{S}}(S_i, d)$ is a sum of independent random variables, the Berry-Esséen bound applies to the probability in (131), and the converse direction of (117) follows since $\|\mathbf{x}\|^2 \leq E$. ∎

If the maximal power constraint in (85) is relaxed to (87), then $E_a^\star(k, d, \epsilon)$, the minimum average power required for transmitting k source samples over an AWGN channel with the probability of exceeding distortion d smaller than or equal to $0 < \epsilon < 1$ satisfies, under assumptions A1–A5:

$$E_a^\star(k, d, \epsilon) \frac{\log e}{N_0} = R(d)(1 - \epsilon)k + O\left(\sqrt{k \log k}\right), \tag{135}$$

i.e. the asymptotically achievable minimum energy per sample is reduced a factor of $1 - \epsilon$ if a maximal power constraint is relaxed to an average one. This parallels the result in (78), which shows that variable-length coding over a channel reduces the asymptotic fundamental limit by a factor of $1 - \epsilon$ compared to fixed-length coding, even without feedback.

Proof of (135): Observe that Theorem 10 ensures that a smaller average energy than that in (135) is not attainable even with full noiseless feedback. In the achievability direction, let $(f^\star, g^\star)$ be the optimal variable-length source code achieving the probability of exceeding d equal to ϵ' (see [5, Section III.B]). Denote by $\ell(f^\star(s))$ the length of $f^\star(s)$. Let M be the size of that code. Set the energy to transmit the codeword of length $\ell(f^\star(S^k))$ to

$$\ell(f^\star(S^k)) N_0 \log_e 2 + \sqrt{k \log k}. \tag{136}$$

As shown in [5], $\mathbb{E}\left[\ell(\mathsf{f}^\star(S))\right]$ is equal to the right side of (72) (with ϵ replaced by ϵ'). Choosing $\epsilon' = \epsilon - \frac{a}{\sqrt{k}}$ for some a, we conclude that indeed the average energy satisfies (135). Moreover, (127) implies that the expression inside the expectation in (100) is $O\left(\frac{1}{\sqrt{k}}\right)$. It follows that for a large enough a, the excess distortion probability is bounded by ϵ. $\blacksquare$

VI. CONCLUSION

We have considered several scenarios for joint source-channel coding with and without feedback. Our main conclusions are:

1) The average delay vs. distortion tradeoff with feedback, as well as the average energy vs. distortion tradeoff with feedback, is governed by the channel capacity, and the source rate-distortion and rate-dispersion functions. In particular, the channel dispersion plays no role.

2) In variable-length coding with feedback, the asymptotically achievable minimum average length is reduced by a factor of $1 - \epsilon$, where ϵ is the excess distortion probability. This asymptotic fundamental limit is approached from below, i.e., counter-intuitively, smaller source blocklengths may lead to smaller attainable average delays.

3) Introducing a termination symbol that is always decoded error-free allows for transmission over noisy channels with guaranteed distortion.

4) Variable-length transmission without feedback still improves the asymptotic fundamental limit by a factor of $1 - \epsilon$, where ϵ is the excess distortion probability.

5) In all the cases we have analyzed the approach to the fundamental limits is very fast: $O\left(\frac{\log k}{k}\right)$, where k is the source blocklength. This behavior is attained, under average distortion, by a separated scheme with stop-feedback.

6) The setting of a wideband Gaussian channel with an energy constraint exhibits many interesting parallels with the variable-length coding setting. Allowing a non-vanishing excess distortion probability ϵ and an average (rather than maximal) energy constraint reduces the asymptotically achievable minimum energy by a factor of $1-\epsilon$. In the presence of feedback, just as in the variable-length coding, this asymptotic fundamental limit is approached from below.

7) Error-free transmission of a random variable W over the AWGN channel with ideal feedback, with almost optimal energy consumption, is possible via a variable-length separated

scheme.

8) More generally, variable-length separated schemes perform remarkably well in all considered scenarios.

REFERENCES

[1] C. E. Shannon, "The zero error capacity of a noisy channel," *IRE Transactions on Information Theory*, vol. 2, no. 3, pp. 8–19, Sep. 1956.

[2] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Problems of Information Transmission*, vol. 12, no. 4, pp. 10–30, 1976.

[3] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, Aug. 2011.

[4] S. L. Fong and V. Y. F. Tan, "Asymptotic expansions for Gaussian channels with feedback under a peak power constraint," *ArXiv preprint arXiv:1410.2390*, Oct. 2014.

[5] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 4316–4330, Aug. 2015.

[6] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2545–2575, May 2013.

[7] A. Tauste Campo, G. Vazquez-Vilar, A. Guillén i Fàbregas, and A. Martinez, "Random-coding joint source-channel bounds," in *Proceedings 2011 IEEE International Symposium on Information Theory*, Saint-Petersburg, Russia, Aug. 2011, pp. 899–902.

[8] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *Proceedings 49th Annual Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sep. 2011.

[9] T. Javidi and A. Goldsmith, "Dynamic joint source-channel coding with feedback," in *IEEE International Symposium on Information Theory*, Istanbul, Turkey, July 2013, pp. 16–20.

[10] G. Caire, S. Shamai, and S. Verdú, "Almost-noiseless joint source-channel coding-decoding of sources with memory," in *Proc. Fifth International ITG Conference on Source and Channel Coding (SCC)*, Jan. 2004, pp. 295–304.

[11] I. A. Ovseevich, "Capacity of a randomized channel with feedback and matching of the source," *Problemy Peredachi Informatsii*, vol. 4, no. 1, pp. 52–59, 1968.

[12] R. Fang, "Optimum linear feedback scheme for matching colored Gaussian sources with white Gaussian channels," *IEEE Transactions on Information Theory*, vol. 16, no. 6, pp. 789–791, 1970.

[13] V. Chande, H. Jafarkhani, and N. Farvardin, "Joint source-channel coding of images for channels with feedback," in *Proceedings 1998 IEEE Information Theory Workshop*, San Diego, CA, Feb. 1998.

[14] V. Kafedziski, "Joint source channel coding of images over frequency selective fading channels with feedback using dct and multicarrier block pulse amplitude modulation," in *Proceedings Asilomar Conf. Signals, Systems and Computers*, vol. 1, 1998, pp. 37–41.

[15] J. Lu, A. Nosratinia, and B. Aazhang, "Progressive joint source-channel coding in feedback channels," in *Proceedings 1999 Data Compression Conference*, 1999, pp. 140–148.

[16] M. Gastpar and B. Rimoldi, "Source-channel communication with feedback," in *Proceedings 2003 IEEE Information Theory Workshop*, Paris, France, 2003, pp. 279–282.

[17] R. Manakkal and B. Rimoldi, "A source-channel coding scheme for discrete memoryless channels with feedback," in *Proceedings 2005 IEEE International Symposium on Information Theory*, Adelaide, Australia, 2005, pp. 1511–1515.

[18] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Transactions on Information Theory*, vol. 9, no. 3, pp. 136–143, 1963.

[19] J. M. Ooi and G. W. Wornell, "Fast iterative coding techniques for feedback channels," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2960–2976, 1998.

[20] R. Ahlswede, "A constructive proof of the coding theorem for discrete memoryless channels with feedback," in *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague, Sep. 1971.

[21] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," *IEEE Transactions on Information Theory*, vol. 57, no. 3, pp. 1186–1222, 2011.

[22] J. Schalkwijk and L. Bluestein, "Transmission of analog waveforms through channels with feedback," *IEEE Transactions on Information Theory*, vol. 13, no. 4, pp. 617–619, 1967.

[23] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[24] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, June 2012.

[25] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: An approach to combined source-channel coding," *IEEE Trans. on Information Theory*, vol. 33, no. 6, pp. 827–838, 1987.

[26] A. Amanullah and M. Salehi, "Joint source-channel coding in the presence of feedback," in *Proc. 27th Asilomar Conf. Sig. Syst. Comp.*, 1993, pp. 930–934.

[27] S. Gorantla and T. Coleman, "Information-theoretic viewpoints on optimal causal coding-decoding problems," *arXiv preprint arXiv:1102.0250*, 2011.

[28] B. Hochwald and K. Zeger, "Tradeoff between source and channel coding," *IEEE Transactions on Information Theory*, vol. 43, no. 5, pp. 1412–1424, 1997.

[29] A. Jain, D. Gunduz, S. R. Kulkarni, H. V. Poor, and S. Verdú, "Energy-distortion tradeoffs in Gaussian joint source-channel coding problems," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3153–3168, 2012.

[30] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Minimum energy to send k bits through the Gaussian channel with and without feedback," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4880–4902, 2011.

[31] S. Verdú and S. Shamai, "Variable-rate channel capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 6, pp. 2651–2667, 2010.

[32] A. Tchamkerten and E. Telatar, "Variable length coding over an unknown channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2126–2145, May 2006.

[33] M. Naghshvar and T. Javidi, "Active sequential hypothesis testing," *The Annals of Statistics*, vol. 41, no. 6, pp. 2703–2738, 2013.

[34] D. Williams, *Probability with Martingales*.   Cambridge University Press, 1991.

[35] A. Wyner, "An upper bound on the entropy series," *Information and Control*, vol. 20, no. 2, pp. 176–181, 1972.

[36] E. C. Posner, E. R. Rodemich, and H. Rumsey, "Epsilon-entropy of stochastic processes," *The Annals of Mathematical Statistics*, vol. 38, no. 4, pp. 1000–1020, 1967.

[37] R. Pilc, "The transmission distortion of a discrete source as a function of the encoding block length," *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 827–885, July/August 1968.

[38] E. Yang and Z. Zhang, "On the redundancy of lossy source coding with abstract alphabets," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1092–1110, May 1999.

[39] T. J. Cruise, "Achievement of rate-distortion bound over additive white noise channel utilizing a noiseless feedback channel," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 583–584, Apr. 1967.

[40] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[41] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*. John Wiley & Sons, 1965.

[42] Y. P. Pyatoshin, "Some properties of m-ary communication systems with coding," *Problemy Peredachi Informatsii*, vol. 4, no. 1, pp. 45–51, 1968.

[43] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Information Theory, Special Issue on Shannon Theory: Perspective, Trends and Applications*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[44] W. Yang, G. Durisi, and Y. Polyanskiy, "Minimum energy to send k bits over Rayleigh-fading channels," in *Proceedings 2015 IEEE International Symposium on Information Theory*, Hong Kong, June 2015, pp. 2959–2963.